Reporting: wragle_report

- Create a **300-600 word written report** called "wrangle_report.pdf" or "wrangle_report.html" that briefly describes your wrangling efforts. This is to be framed as an internal document.

This report describe effort on WeRateDog data set, i used three different set of data downloaded programatically from different sources. the enitere project was completed on udacity project work space, but due to some qaulity and tidines issues on the both downloaded data set, i apply three wrangling process in order to make my data set clean and ready for any analysis.

1.Gathering Data 2.Assessing Data 3.Data cleaning

1. Gathering Data

I gather all my data from three dofferent sources a. Enhanced twitter archive data set, it contain all the basic tweet data

b. Twitter API data set, it contain some omission column on enhanced archieve data set, retweet count and favorite count

c. Image prediction data set, it contain all the image prediction after using nueral network to classfied breeds of dog

2. Assessing Data

After Data gathering the next stage process on data wrangling is assessing my data set using both visual and programatical method in order to identified if there is any data qaulity and tidnes issues. the folloeing data qaulity and tidnes issues where identified

Quality issues Archive twitter data set

1.Missing Data

2.Tweet_id it is not in String data type format

3.timestamp it is not in datetime format

4.Some invalid dog names

5.Some of Denominator rating are lessthan 10 while some exceed 10, it's suppose to be 10

6.some numerator rating are incorrect i.e 1776, 666 and so on. it's suppose to be from 1 to 15

7.Some p1,p2 and p3 name used underscore

Image prediction data set

8.Some P1,P2,P3 first latter start with small latter while other start with capital latter

9.Missing Data 2075 instead of 2354

Tidiness issues 1.All data have related with each other but seperated differently

2.Dog breeds column is separated into four different column

3. Data cleanig

After assessing all my data set the next stage process on data wrangling is clean all of the issues i documented while assessing.

Finally my data set is clean and ready for analysis. I saved the master table to twitter_archive_master.csv